GDIT

Date: June 30, 2021 Docket No. OP-1743

FEDERAL RESERVE

REQUEST FOR INFORMATION AND COMMENT ON FINANCIAL INSTITUTIONS' USE OF ARTIFICIAL INTELLIGENCE, INCLUDING MACHINE LEARNING

Submitted Electronically to:

Federal Reserve

regs.comments@federalreserve.gov

Submitted by:

General Dynamics Information Technology, Inc.

POC Name: Frederick Stith, Contracts Lead

Address: 3150 Fairview Park Drive, Suite 100

Falls Church, VA 22042-4504

Phone: (667) 786-7843

Email: Frederick.stith@gdit.com

www.gdit.com

DISCLOSURE STATEMENT

This RFI includes data that shall not be disclosed outside the Government and shall not be duplicated, used, or disclosed—in whole or in part—for any purpose other than to evaluate this RFI. This restriction does not limit the Government's right to use information contained in this data if it is obtained from another source without restriction. The data subject to the restriction are contained in sheets with the following legend: "Use or disclosure of data contained on this sheet is subject to the restriction on the title page of this RFI Response"

GDIT Introduction

General Dynamics Information Technology, Inc. (GDIT) is a Federal System Integrator (FSI) and global leader, providing full-lifecycle information technology (IT) services that include leading solutions and technologies leveraging Artificial Intelligence (AI) and Machine Learning (ML). GDIT understands the challenges with regards to the utilization of AI by financial institutions in their provision of services to customers and for other business or operational purposes; appropriate governance, risk management, and controls over AI; and any challenges in developing, adopting, and managing AI.

GDIT looks forward to our continued support as a trusted advisor and leading industry expert to facilitate financial institutions' use of AI in a safe and sound manner and in compliance with applicable laws and regulations, including those related to consumer protection.

GDIT Response

Question:

> ***How do financial institutions identify and manage risks relating to AI Explainability?***

Explainability, in the context of an AI model making an inference, is the ability for a human to understand why a particular conclusion was reached.

Risk, in this context, is when a model unfairly makes an inference that disadvantages a group or cohort – but cannot be explained, nor understood by a human.

> ***What barriers or challenges for Explainability exist for developing, adopting, and managing AI?***

AI models are quite difficult to interpret. An examination of the model architecture can provide insights into how a decision was made for some methods, but not all. Black-box methods (Neural networks, Deep Learning Networks) are especially difficult to interprets.

This barrier is being overcome by explainable AI methods (e.g., Local Interpretable Model-Agnostic Explanation (LIME), Shapley) that offer insights into the underlying model components. Interpretability frameworks (e.g., Saliency, Integrated Gradients, Occlusion) highlight the key features used to make a classification by deep learning-based models are beginning to be used to explain Black-Box methods.

Question:

> ***How do financial institutions use post-hoc methods to assist in evaluating conceptual soundness? How common are these methods?***

Post hoc methods are used to explain why a decision was chosen over alternative decisions. Model inputs are varied to analyze the changes to outcomes and the result is a set of inferences as to why a path was likely chosen. The approach is flexible, model agnostic and often provides Explainability to the consumer. LIME and Shapley methods are gaining use across wide set of applications. In addition, Interpretability Frameworks (Saliency, Integrated Gradients, Occlusion) are especially useful in deep learning classification inferences.

Are there limitations of these methods (whether to explain an AI approach's overall operation or to explain a specific prediction or categorization)? If so, please provide details on such limitations.

Explainability is an emerging area, with NIST standards and research into methods beginning to show promising results. Explainability frameworks are specific to each algorithmic method, and hence require numerous implementations and complexity in presenting to end users, be they data scientists testing and evaluating a model, or an end-user looking for support for an inference.

Conversely - the model agnostic approach is a generic approach that makes inferences made on varied input (and output) from the model rather than factual understanding of the model itself. It cannot be known with certainty whether the inferences are correct, whether the varied input caused the new outputs, and, often, it is not possible to test all variations of the input data combinations or provide full inferences across a complex input set. Further, while inputs are known, the range of the inputs, and the range of the outputs are often unknown leading to limitations within the approach.

Question:

For which uses of AI is lack of Explainability more of a challenge? Please describe those challenges in detail. How do financial institutions account for and manage the varied challenges and risks posed by different uses?

Explainability is critical in cases when the use of AI impacts businesses or consumers, perhaps by denying a service (e.g., a loan) or recommending a treatment that can be interpreted as unfair when compared to others (e.g., a credit score). To be trusted, AI must be both explainable and accountable – providing reasons for adverse outcomes and supporting redress should an unfair decision be rendered. . In these cases, the inability to provide sufficiently detailed reasons lead to individual or even class litigation (and possible liability) in either civil or criminal forums as well as regulatory actions.

In contrast, AI that governs internal, non-regulated processes often has lower requirements for Explainability, since the business simply prioritizes a favorable outcome rather than the Explainability of process that resulted in the outcome. Examples of these AI applications may include AIOps to maintain infrastructure, help desk automation, customer contact automation, and marketing and sales campaigns. Although diligence with respect to bias in any customer setting is fundamental best practice.

How do financial institutions account for and manage the varied challenges and risks posed by different uses?

Financial institutions use policies and procedures to guide what AI may impact a party adversely or is regulated, then apply more stringent (and costly) guidelines on the use of AI in such cases. One emerging trend is the concept of an AI Review board, a governing body that reviews the mission and the proposed solution, assesses risk and defines proactive data bias and model bias statistical monitoring requirements. Further, AI boards also designate responsible parties, and avenues for redress should the AI cause harm.

Question:

How do financial institutions using AI manage risks related to data quality and data processing?

Financial institutions institute data management policies and procedures, which define the quality standards throughout the data lifecycle. The policies typically cover seven characteristics of data quality: Consistency, Accuracy, Completeness, Auditability, Validity, Uniqueness, Timeliness.

Further, since AI models are built on data, that data must examined for bias, and if present, sampled in ways that balance the data before training AI models. While in operation, the data must be monitored for drift away from the statistical characteristics that it embodied when training the original AI models. If not, the new data may make incorrect, or unfair decisions until the models can be re-trained with new data.

How, if at all, have control processes or automated data quality routines changed to address the data quality needs of AI?

Financial institutions house data in data lake and data warehouses to support AI capabilities. Data loaded from Line of Business (LOB) systems into big data solutions through various data stages: landing data, transforming the data into a clean layer, and finally into a curated layer.

Data quality is monitored throughout the stages by measuring the data at each stage against the seven characteristics as identified by data stewards. What institutions chose to do with data that fails the quality standards varies. Some institutions merely measure the data quality at each stage and report on it to leadership. Others identify "data gates" wherein data that fails the quality checks at each stage is not allowed to progress to the next stage. Still others have defined data quality checks at the source system level.

DataOps is having the same profound impact on data management, pipelines, and data products as DevSecOps had on code development and service deployment. In DataOps, data is the input, data products as the final products, data pipeline version as input, and a deployed pipeline as the product. Quality checks are incorporated into the pipelines themselves and pipeline development. Versioning throughout the data ecosystem from schema versioning to versioning of the data itself. In short, DataOps provides the ability to roll pipeline code changes back automatically and perform rollbacks on the data itself should either fail data tests. In summary, versioning across schemas, pipelines, services, and data itself is allowing a robust DataOps environment with automated deployments, testing, and, as necessary, rollbacks.

In the case of AI, data quality must be extended into the testing and mitigation of data bias. Data bias may occur within selection of training, testing data, feature selection, and more. Similar to DataOps, MLOps is an automated framework with similar capabilities for integrating the training, testing (unit, AI tests), deployment, and, if required, rollback.

How does risk management for alternative data compare to that of traditional data?

Alternative data may be intentionally or unintentionally tainted by external parties. As such, additional risk management is needed to manage risk based off the determined authority of the producer and content.

Are there any barriers or challenges that data quality and data processing pose for developing, adopting, and managing AI? If so, please provide details on those barriers or challenges.

There are significant data challenges to developing AI. First and foremost, is collecting of a large set of labeled data that is representative of the various outcomes to be predicted as well as the possible combinations of input variables. Machine learning algorithms need a substantially large data set on which to train. The second challenge is data structure. Most data is collected from systems designed for another purpose. This data must be "shaped" by data engineers through transformations, enrichments, and

cleaning. This data preparation effort still consumes up to 80% of a data scientists' effort – though new tools are emerging to speed this task.

Question:

Are there specific uses of AI for which alternative data are particularly effective?

Alternative data has numerous effective uses within financial institutions. This data includes text (press releases, news, social media, filings, etc.) and numerical data (stock indices, futures, foreign markets, etc.). These sources are used to fulfil numerous AI uses including:

Creditworthiness Assessment: Using cashflow markers, academic, and work history

Financial Markets: Using press releases by bellwethers will swing markets, News reports by agencies, Metric releases such as manufacturing indexes, Social activity by influences predicts short to long term trends in share prices impacting various downstream financial impacts, to social media activity predicting network size. In a notable example, twitter mood has correlations to stock price, behavior, future paths, and the Dow Jones itself. Specifically, twitter allows the integration of behavioral finance into models.

Real-Time Financial Market: Using real-time market input from the various sectors combined with alternative data (economic, "real" economy, social media sentiment, media sentiment) an AI could recommend regulatory or correction action and allow the action to be taken in near real-time.

Economic or Corporate Activity Measurement: Using satellite photos or drone imagery, one can determine the output of a car company and delivery rates of inventory months prior to the results being announced thus allowing understanding of (un)favorability and prediction of quarterly financial results including revenue, profitability, expenses. This has a material impact on numerous financial segments from trading, risk, and financing. Analysis of training grounds, Twitter, and other sources will often pick up rumors (along with probability) of new models being released. Public sentiment can be used as input for models with predictions of sale volumes, revenue, etc.

Estimates: Aggregations of the above may be used to improve predictions and modeling of key economic measurements such as future GDP allowing adjustments to present environment BEFORE the impact occurs on real GDP.

Question:

How do financial institutions manage AI risks relating to overfitting?

Institutions manage overfitting risk during the MLOps process by measuring the error rates, variances and performing k-fold cross-validation. Overfitting is managed by properly stratifying and sampling training and test data sets, ensuring the model doesn't train too long on sample data (early stopping), feature selection, etc. MLOps helps by automating the performance of many of these activities, assisting in avoiding overfitting to begin with, and identifying overfitting when it does occur.

What barriers or challenges, if any, does overfitting pose for developing, adopting, and managing AI?

Overfitting prevents a model from properly generalizing against unknown or unseen populations. In short, an overfitting model will not accurately provide predictions when it encounters unseen / new populations

or new inputs in general. An overfitted model is unreliable and thus poses risk when presented with data unlike that on which it was trained.

How do financial institutions develop their AI so that it will adapt to new and potentially different populations (outside of the test and training data)?

Financial institutions want to avoid over or underfitting the data. A good model that doesn't over or underfit will generalize well for new populations that are not covered by training data. The model should be trained to the extent that a good bias-variance tradeoff exists.

Further, model performance is monitored to determine if its accuracy is declining, most likely caused by data drift. When this occurs, a MLOPs pipeline with more recent data is invoked to retrain, test, evaluate, and deploy the model. In some organizations, this CI/CD process is 100% automated, with new models published rapidly with great frequency.

Question:

Have financial institutions identified particular cybersecurity risks or experienced such incidents with respect to AI? If so, what practices are financial institutions using to manage cybersecurity risks related to AI?

Cybersecurity risks exist to models, how they're trained/tested, feature selection, data, pipelines, data quality, and more. Cybersecurity risks from AI exist throughout data management -- from processing pipelines, quality measures, to the models themselves. The security in these areas fall within the typical infrastructure, code, and data security measures one would normally put in place from data quality testing & monitoring, data quality unit tests, data quality gates, model testing and monitoring, tokenized security, IAM roles, reducing surface area, firewalls, shields, etc.

Please describe any barriers or challenges to the use of AI associated with cybersecurity risks.

Beyond the expected threats above, alternative data poses a cybersecurity risk in that it may be intentionally manipulated to manipulate models and could dramatically effect performance. Some AI models are susceptible to "data poisoning" in which bad data is maliciously sent to a model endpoint for inference. The data stream may overwhelm the endpoint with volume and deny service to legitimate users. Or the malicious stream may cause inferences that sow confusion, consume resources, and require undue attention to repair.

Are there specific information security or cybersecurity controls that can be applied to AI?

This means of attack might be intercepted with data quality measures to review data before running the models.

Question:

How do financial institutions manage AI risks relating to dynamic updating? Describe any barriers or challenges that may impede the use of AI that involve dynamic updating. How do financial institutions gain an understanding of whether AI approaches producing different outputs over time based on the same inputs are operating as intended?

AI Models are trained on a labeled data set, collected at point in time. If the external world changes and new data characteristics emerge, the models may exhibit a reduced accuracy and need to be retrained, tested, and redeployed. This is one form of dynamic updating. It is controlled and managed because data is run through the same MLOps pipeline, which means the new variation is assessed against the original quality standards, goal posts, and tests of the any other ML model including the possibility of model rejection / failure, rollback if variances are discovered, and human intervention.

New models may, and often do, produce different outputs based on the same inputs. This reflects the state of an evolving external world or system. The understanding is that this is a new model built on new data. It is operating as intended.

Question:

> ***Please describe any particular challenges or impediments financial institutions face in using AI developed or provided by third parties and a description of how financial institutions manage the associated risks. Please provide detail on any challenges or impediments. How do those challenges or impediments vary by financial institution size and complexity?***

Third-party AI are essentially a short-cut that allows financial institutions to bypass much of the DataOps and MLOps activities. If done carefully, these pre-built solutions can be integrated into the overall framework if they can meet policies and procedures for model, testing, and monitoring as well as data preparation and verification.

Question:

> ***What techniques are available to facilitate or evaluate the compliance of AI-based credit determination approaches with fair lending laws or mitigate risks of non-compliance? Please explain these techniques and their objectives, limitations of those techniques, and how those techniques relate to fair lending legal requirements.***

Model testing, monitoring, and Explainability can be used to show compliance with appropriate laws. Model Explainability will show what considerations went into the models and the reasoning for the results. The combination gives institutions and impacted businesses or consumers visibility into how outcomes were determined and can be used to ensure compliance with fair lending laws.

If not in compliance, Explainability can provide impacted parties with paths to understand and remediate the driving factors for adverse actions and provide institutions with the ability to understand model decisions, impacted party's unique situations, and how those situations impacted the decision. As such, it gives the impacted parties enough information to understand and remediate the causes or to appeal the action.

Question:

> ***What are the risks that AI can be biased and/or result in discrimination on prohibited bases?***

The risks that AI can be biased is determined by the underlying labeled data used to train the models. Many historical data sets record determinations and judgements made by humans in a social context. From real estate red lines to lending decisions to credit rating to insurance underwriting, our collective

history is fraught decisions that were made in a discriminatory manner. While great progress driven by laws and regulations have improved this social condition, that bias may still be present in training data.

Are there effective ways to reduce risk of discrimination, whether during development, validation, revision, and/or use?

Identifying and removing bias through statistical sampling can reduce this impact. This approach should also identify bias in a feature proxy that is highly correlated to the biased data element (e.g. zip code as a social determinant of health outcomes and wealth).

Institutions are increasingly using robust model unit testing, validation testing, and monitoring to independently check model outcomes (including for discrimination and compliance) against quality guidelines which include discrimination "guard rails". Model Accelerators provide the ability to review multiple potential models with inclusivity as a consideration and these would be embedded within the MLOps pipeline. Additionally, institutions will use MLOps pipelines to determine the training/testing data automatically for new models and this automated selection will reduce the potential for bias further.

What are some of the barriers to or limitations of those methods?

Like other topics, an ample supply of labeled data is critical to identifying and removing bias.

Question:

To what extent do model risk management principles and practices aid or inhibit evaluations of AI-based credit determination approaches for compliance with fair lending laws?

Model risk management is embedded within MLOps and a standard part of model testing. We consider model risk management to be complimentary and an accelerator of good model development and lifecycles. Institutions are integrating model risk management and compliance into the model testing/validation as previously mentioned.

Question:

As part of their compliance management systems, financial institutions may conduct fair lending risk assessments by using models designed to evaluate fair lending risks ("fair lending risk assessment models"). What challenges, if any, do financial institutions face when applying internal model risk management principles and practices to the development, validation, or use of fair lending risk assessment models based on AI?

Like other topics, an ample supply of labeled data is critical to identifying and removing bias. If financial institutions have examples of fair AND unfair lending, then a model could be developed. Can a large collection of unfair lending examples can be secured? If not, this approach is unworkable.

Question:

The Equal Credit Opportunity Act (ECOA), which is implemented by Regulation B, requires creditors to notify an applicant of the principal reasons for taking adverse action for credit or to provide an applicant a disclosure of the right to request those reasons. What approaches

can be used to identify the reasons for taking adverse action on a credit application, when AI is employed? Does Regulation B provide sufficient clarity for the statement of reasons for adverse action when AI is used? If not, please describe in detail any opportunities for clarity.

While emergent, Model Agnostic Explainability frameworks such as LIME and Shapley should be able to provide sufficient details for why the adverse reaction occurred regardless of what model is used.

In addition, algorithm specific explicit Explainability methods can be used to pinpoint reasons for adverse event determinations.

Question:

To the extent not already discussed, please identify any additional uses of AI by financial institutions and any risk management challenges or other factors that may impede adoption and use of AI.

Financial institutions can apply AI to many regulatory missions, as well as IT operations, Customer Support, and Intelligent Automation. Some notable missions include:

- predict inflation across a wide spectrum of market and economic data sources
- predict manufacturing activity, capabilities, in advance of surveys
- forecast the impact of rate increases on the economy, markets, jobs
- predict unemployment, job growth, and related metrics
- predict cost of living, health outcomes, and related quality of life metrics

Question:

To the extent not already discussed, please identify any benefits or risks to financial institutions' customers or prospective customers from the use of AI by those financial institutions. Please provide any suggestions on how to maximize benefits or address any identified risks.

Significant risks exist around risk of biased data being supplied into the models. Bias may not be readily apparent and may be introduced unintentionally. Data bias detection and model outcome analysis are keys to reducing this risk.

AI has the capability to remove bias from the system, processes, and outcomes. When the lifecycle is properly managed, everyone gets a fair opportunity to have positive outcomes, on a level playing field, and, when adverse events happen, everyone understands why those events occurred, how the decisions were arrived at, what the model considered, and what they can do to change contributing factors or appeal the outcome.